REGULARITY OF THE CONDITIONAL EXPECTATIONS WITH RESPECT TO SIGNAL TO NOISE RATIO

A. S. ÜSTÜNEL
TOT PASSA, PERÒ SEGUEIX SENT L'AMISTAT

Abstract: Let (W, H, μ) be the classical Wiener space, assume that $U_\lambda = I_W + u_\lambda$ is an adapted perturbation of identity where the perturbation u_λ is an H-valued map, defined up to μ-equivalence classes, such that its Lebesgue density $s \to \dot{u}_\lambda(s)$ is almost surely adapted to the canonical filtration of the Wiener space and depending measurably on a real parameter λ. Assuming some regularity for u_λ, its Sobolev derivative and integrability of the divergence of the resolvent operator of its Sobolev derivative, we prove the almost sure and L^p-regularity w.r. to λ of the estimation $E[\dot{u}_\lambda(s)|\mathcal{U}_\lambda(s)]$ and more generally of the conditional expectations of the type $E[F \mid \mathcal{U}_\lambda(s)]$ for nice Wiener functionals, where $(\mathcal{U}_\lambda(s), s \in [0,1])$ is the the filtration which is generated by U_λ. These results are applied to prove the invertibility of the adapted perturbations of identity, hence to prove the strong existence and uniqueness of functional SDE's; convexity of the entropy and the quadratic estimation error and finally to the information theory.

Contents

Keywords: Entropy, adapted perturbation of identity, Wiener measure, invertibility.

1. Introduction

The Malliavin calculus studies the regularity of the laws of the random variables (functionals) defined on a Winer space (abstract or classical) with values in finite dimensional Euclidean spaces (more generally manifolds) using a variational calculus in the direction of the underlying quasi-invariance space, called the Cameron-Martin space. Although its efficiency is globally recognized by now, for

the maps taking values in the infinite dimensional spaces the Malliavin calculus does not apply as easily as in the finite dimensional case due to the absence of the Lebesgue measure and even the problem itself needs to be defined. For instance, there is a notion called signal to noise ratio which finds its roots in engineering which requires regularity of infinite dimensional objects with respect to finite dimensional parameters (cf.[1, 8, 9, 10, 11, 12]). Let us explain the problem along its general lines briefly: imagine a communication channel of the form $y = \sqrt{\lambda}x + w$, where x denotes the emitted signal and w is a noise which corrupts the communications. The problem of estimation of the signal x from the data generated y is studied since the early beginnings of the electrical engineering. One of the main problems dealt with is the behavior of the L^2-error of the estimation w.r. to the signal to noise ratio λ. This requires elementary probability when x and w are independent finite dimensional variables, though it gives important results for engineers. In particular, it has been recently realized that (cf. [8, 22]), in this linear model with w being Gaussian, the derivative of the mutual information between x and y w.r. to λ equals to the half of the mean quadratic error of estimation. The infinite dimensional case is more tricky and requires already the techniques of Wiener space analysis and the Malliavin calculus (cf. [22]). The situation is much more complicated in the case where the signal is correlated to the noise; in fact we need the λ-regularity of the conditional expectations w. r. to the filtration generated by y, which is, at first sight, clearly outside the scope of the Malliavin calculus.

In this paper we study the generalization of the problem mentioned above. Namely assume that we are given, in the setting of a classical Wiener space, denoted as (W, H, μ), a signal which is of the form of an adapted perturbation of identity:

$$U_\lambda(t, w) = W_t(w) + \int_0^t \dot{u}_\lambda(s, w)ds\,,$$

where $(W_t, t \in [0, 1])$ is the canonical Wiener process, $\dot{u}_\lambda$ is an element of $L^2(ds \times d\mu)$ which is adapted to the Brownian filtration ds-almost surely and λ is a real parameter. Let $\mathcal{U}_\lambda(t)$ be the sigma algebra generated by $(U_\lambda(s),\ s \leq t)$. What can we say about the regularity, i.e., continuity and/or differentiability w.r. to λ, of the functionals of the form $\lambda \to E[F \mid \mathcal{U}_\lambda(t)]$ and $\lambda \to E[F \mid U_\lambda = w]$ (the latter denotes the disintegration) given various regularity assumptions about the map $\lambda \to \dot{u}_\lambda$, like differentiability of it or its H-Sobolev derivatives w.r. to λ? We prove that the answer to these questions depend essentially on the behavior of the random resolvent operator $(I_H + \nabla u_\lambda)^{-1}$, where ∇u_λ denotes the Sobolev derivative of u_λ, which is a quasi-nilpotent Hilbert-Schmidt operator, hence its resolvent exists always. More precisely we prove that if the functional

$$(1 + \rho(-\delta u_\lambda)\delta\left((I_H + \nabla u_\lambda)^{-1}\frac{d}{d\lambda}u_\lambda\right) \tag{1.1}$$

is in $L^1(d\lambda \times d\mu, [0, M] \times W)$ for some $M > 0$, where δ denotes the Gaussian divergence and $\rho(-\delta u)$ is the Girsanov-Wick exponential corresponding to the stochastic integral δu (cf. the next section), then the map $\lambda \to L_\lambda$ is absolutely continuous almost surely where L_λ is the Radon-Nikodym derivative of $U_\lambda\mu$ w.r. to μ and we can calculate its derivative explicitly. This observation follows from some variational calculus and from the Malliavin calculus. The iteration of the hypothesis (1.1) by replacing $\delta((I_H + \nabla u_\lambda)^{-1}\frac{d}{d\lambda}u_\lambda)$ with its λ-derivatives permits us to prove the higher order differentiability of the above conditional expectations w.r. to λ and these results are exposed in Section 3. In Section 4, we give applications of these results to show the almost sure invertibility of

the adapted perturbations of the identity, which is equivalent to the strong existence and uniqueness results of the (functional) stochastic differential equations. In Section 5, we apply the results of Section 3 to calculate the derivatives of the relative entropy of $U_\lambda \mu$ w.r. to μ in the general case, i.e., we do not suppose the a.s. invertibility of U_λ, which demands the calculation of the derivatives of the non-trivial conditional expectations. Some results are also given for the derivative of the quadratic error in the case of anticipative estimation as well as the relations to the Monge-Kantorovich measure transportation theory and the Monge-Ampère equation. In Section 6, we generalize the celebrated result about the relation between the mutual information and the mean quadratic error (cf. [1, 9, 10]) in the following way: we suppress the hypothesis of independence between the signal and the noise as well as the almost sure invertibility of the observation for fixed exterior parameter of the signal. With the help of the results of Section 3, the calculations of the first and second order derivatives of the mutual information w.r. to the ratio parameter λ are also given.

2. **Preliminaries and notation**

Let W be the classical Wiener space $C([0,T],\mathbb{R}^n)$ with the Wiener measure μ. The corresponding Cameron-Martin space is denoted by H. Recall that the injection $H \hookrightarrow W$ is compact and its adjoint is the natural injection $W^\star \hookrightarrow H^\star \subset L^2(\mu)$. Since the image of μ under the mappings $w \to w+h$, $h \in H$ is equivalent to μ, the Gâteaux derivative in the H direction of the random variables is a closable operator on $L^p(\mu)$-spaces and this closure is denoted by ∇ and called the Sobolev derivative (on the Wiener space) cf., for example [13, 14]. The corresponding Sobolev spaces consisting of (the equivalence classes) of real-valued random variables will be denoted as $\mathbb{D}_{p,k}$, where $k \in \mathbb{N}$ is the order of differentiability and $p>1$ is the order of integrability. If the random variables are with values in some separable Hilbert space, say Φ, then we shall define similarly the corresponding Sobolev spaces and they are denoted as $\mathbb{D}_{p,k}(\Phi)$, $p>1$, $k \in \mathbb{N}$. Since $\nabla : \mathbb{D}_{p,k} \to \mathbb{D}_{p,k-1}(H)$ is a continuous and linear operator its adjoint is a well-defined operator which we represent by δ. A very important feature in the theory is that δ coincides with the Itô integral of the Lebesgue density of the adapted elements of $\mathbb{D}_{p,k}(H)$ (cf.[13, 14]).

For any $t \ge 0$ and measurable $f : W \to \mathbb{R}_+$, we note by

$$P_t f(x) = \int_W f\left(e^{-t}x + \sqrt{1-e^{-2t}}y\right)\mu(dy)\,,$$

it is well-known that $(P_t, t \in \mathbb{R}_+)$ is a hypercontractive semigroup on $L^p(\mu), p>1$, which is called the Ornstein-Uhlenbeck semigroup (cf.[13, 14]). Its infinitesimal generator is denoted by $-\mathcal{L}$ and we call $\mathcal{L}$ the Ornstein-Uhlenbeck operator (sometimes called the number operator by the physicists). The norms defined by

$$\|\phi\|_{p,k} = \|(I+\mathcal{L})^{k/2}\phi\|_{L^p(\mu)} \tag{2.2}$$

are equivalent to the norms defined by the iterates of the Sobolev derivative ∇. This observation permits us to identify the duals of the space $\mathbb{D}_{p,k}(\Phi); p>1$, $k \in \mathbb{N}$ by $\mathbb{D}_{q,-k}(\Phi')$, with $q^{-1} = 1-p^{-1}$, where the latter space is defined by replacing k in (2.2) by $-k$, this gives us the distribution spaces on the Wiener space W (in fact we can take as k any real number). An easy calculation shows that, formally, $\delta \circ \nabla = \mathcal{L}$, and this permits us to extend the divergence and the derivative operators to the distributions as linear, continuous operators. In fact $\delta : \mathbb{D}_{q,k}(H \otimes \Phi) \to \mathbb{D}_{q,k-1}(\Phi)$ and $\nabla : \mathbb{D}_{q,k}(\Phi) \to \mathbb{D}_{q,k-1}(H \otimes \Phi)$ continuously, for any $q>1$ and $k \in \mathbb{R}$, where $H \otimes \Phi$ denotes the

completed Hilbert-Schmidt tensor product (cf., for instance [13, 14, 19]). We shall denote by $\mathbb{D}(\Phi)$ and $\mathbb{D}'(\Phi)$ respectively the sets

$$\mathbb{D}(\Phi) = \bigcap_{p>1, k\in\mathbb{N}} \mathbb{D}_{p,k}(\Phi)\,,$$

and

$$\mathbb{D}'(\Phi) = \bigcup_{p>1, k\in\mathbb{N}} \mathbb{D}_{p,-k}(\Phi)\,,$$

where the former is equipped with the projective and the latter is equipped with the inductive limit topologies.

Let us denote by $(W_t, t \in [0,1])$ the coordinate map on W which is the canonical Brownian motion (or Wiener process) under the Wiener measure, let $(\mathcal{F}_t, t \in [0,1])$ be its completed filtration. The elements of $L^2(\mu, H) = \mathbb{D}_{2,0}(H)$ such that $w \to \dot{u}(s,w)$ are ds-a.s. $\mathcal{F}_S$ measurable will be noted as $L^2_a(\mu, H)$ or $\mathbb{D}^a_{2,0}(H)$. $L^0_a(\mu, H)$ is defined similarly (under the convergence in probability). Let $U : W \to W$ be defined as $U = I_W + u$ with some $u \in L^0_a(\mu, H)$, we say that U is μ-almost surely invertible if there exists some $V : W \to W$ such that $V\mu \ll \mu$ and that

$$\mu\left\{w : U \circ V(w) = V \circ U(w) = w\right\} = 1\,.$$

The following results are proved with various extensions in [15, 16, 17]:

Theorem 1. *Assume that $u \in L^0_a(\mu, H)$, let L be the Radon-Nikodym density of $U\mu = (I_W + u)\mu$ w.r. to μ, where $U\mu$ denotes the image (push forward) of μ under the map U. Then we have*

(1)
$$E[L \log L] \leq \frac{1}{2}\|u\|^2_{L^2(\mu,H)} = \frac{1}{2}E\int_0^1 |\dot{u}_s|^2 ds\,.$$

(2) *Assume that $E[\rho(-\delta u)] = 1$, then we have the equality:*

$$E[L \log L] = \frac{1}{2}\|u\|^2_{L^2(\mu,H)} \tag{2.3}$$

if and only if U is almost surely invertible and its inverse can be written as $V = I_W + v$, with $v \in L^0_a(\mu, H)$.

(3) *Assume that $E[L \log L - \log L] < \infty$ and the equality (2.3) holds, then U is again almost surely invertible and its inverse can be written as $V = I_W + v$, with $v \in L^0_a(\mu, H)$.*

The following result gives the relation between the entropy and the estimation (cf. [15] for the proof):

Theorem 2. *Assume that $u \in L^2_a(\mu, H)$, let L be the Radon-Nikodym density of $U\mu = (I_W + u)\mu$ w.r. to μ, where $U\mu$ denotes the image (push forward) of μ under the map U and let $(\mathcal{U}_t, t \in [0,1])$ be the filtration generated by $(t,w) \to U(t,w)$. Assume that $E[\rho(-\delta u)] = 1$. Then we have*

- $$E[L \log L] = \frac{1}{2}E\int_0^1 |E[\dot{u}_s \mid \mathcal{U}_s]|^2 ds\,.$$
- $$L \circ U\, E[\rho(-\delta u)|U] = 1$$

  μ-almost surely.

3. Basic results

Let (W,H,μ) be the classical Wiener space, i.e., $W=C_0([0,1],\mathbb{R}^d)$, $H=H^1([0,1],\mathbb{R}^d)$ and μ is the Wiener measure under which the evaluation map at $t\in[0,1]$ is a Brownian motion. Assume that $U_\lambda:W\to W$ is defined as

$$U_\lambda(t,w)=W_t(w)+\int_0^t \dot u_\lambda(s,w)ds\,,$$

with $\lambda\in\mathbb{R}$ being a parameter. We assume that $\dot u_\lambda\in L^2_a([0,1]\times W,dt\times d\mu)$, where the subscript "$_a$" means that it is adapted to the canonical filtration for almost all $s\in[0,1]$. We denote the primitive of $\dot u_\lambda$ by u_λ and assume that $E[\rho(-\delta u_\lambda)]=1$, where ρ denotes the Girsanov exponential:

$$\rho(-\delta u_\lambda)=\exp\left(-\int_0^1 \dot u_\lambda(s)dW_s-\frac12\int_0^1|\dot u_\lambda(s)|^2ds\right)\,.$$

We shall assume that the map $\lambda\to\dot u_\lambda$ is differentiable as a map in $L^2_a([0,1]\times W,dt\times d\mu)$, we denote its derivative w.r. to λ by $\dot u'_\lambda(s)$ or by $\dot u'(\lambda,s)$ and its primitive w.r. to s is denoted as $u'_\lambda(t)$.

Theorem 3. *Suppose that $\lambda\to u_\lambda\in L^p_{loc}(\mathbb{R},d\lambda;\mathbb{D}_{p,1}(H))$ for some $p\geq1$, with $E[\rho(-\delta u_\lambda)]=1$ for any $\lambda\geq0$ and also that*

$$E\int_0^\lambda(1+\rho(-\delta u_\alpha))\,|E[\delta(K_\alpha u'_\alpha)|U_\alpha]|^p\,d\alpha<\infty\,,$$

where $K_\alpha=(I_H+\nabla u_\alpha)^{-1}$. Then the map

$$\lambda\to L_\lambda=\frac{dU_\lambda\mu}{d\mu}$$

is absolutely continuous and we have

$$L_\lambda(w)=L_0\exp\int_0^\lambda E\Big[\delta(K_\alpha u'_\alpha)|U_\alpha=w\Big]d\alpha\,.$$

Proof: Let us note first that the map $(\lambda,w)\to L_\lambda(w)$ is measurable thanks to the Radon-Nikodym theorem. Besides, for any (smooth) cylindrical function f, we have

$$\begin{aligned}\frac{d}{d\lambda}E[f\circ U_\lambda] &= E[(\nabla f\circ U_\lambda,u'_\lambda)_H]\\ &= E[((I_H+\nabla u_\lambda)^{-1\star}\nabla(f\circ U_\lambda),u'_\lambda)_H]\\ &= E[(\nabla(f\circ U_\lambda),(I_H+\nabla u_\lambda)^{-1}u'_\lambda)_H]\\ &= E[f\circ U_\lambda\,\delta\{(I_H+\nabla u_\lambda)^{-1}u'_\lambda\}]\\ &= E[f\circ U_\lambda\,E[\delta(K_\lambda u'_\lambda)|U_\lambda]]\\ &= E[f\,E[\delta(K_\lambda u'_\lambda)|U_\lambda=w]L_\lambda]\,.\end{aligned}$$

Hence, for any fixed f, we get

$$\frac{d}{d\lambda}\langle f,L_\lambda\rangle=\langle f,L_\lambda E[\delta(K_\lambda u'_\lambda)|U_\lambda=w]\rangle\,,$$

both sides of the above equality are continuous w.r. to λ, hence we get

$$<f,L_\lambda>-<f,L_0>=\int_0^\lambda<f,L_\alpha E\left[\delta(K_\alpha u'_\alpha)|U_\alpha=w\right]>d\alpha\,.$$

From the hypothesis, we have

$$E\int_0^\lambda L_\alpha |E[\delta(K_\alpha u'_\alpha)|U_\alpha = w]|d\alpha = E\int_0^\lambda |E[\delta(K_\alpha u'_\alpha)|U_\alpha]|\,d\alpha < \infty\,.$$

By the measurability of the disintegrations, the mapping $(\alpha, w) \to E[\delta(K_\alpha u'_\alpha)|U_\alpha = w]$ has a measurable modification, hence the following integral equation holds in the ordinary sense for almost all $w \in W$

$$L_\lambda = L_0 + \int_0^\lambda L_\alpha E[\delta(K_\alpha u'_\alpha)|U_\alpha = w]d\alpha\,,$$

for $\lambda > 0$. Therefore the map $\lambda \to L_\lambda$ is almost surely absolutely continuous w.r. to the Lebesgue measure. To show its representation as an exponential, we need to show that the map $\alpha \to E[\delta(K_\alpha u'_\alpha)|U_\alpha = w]$ is almost surely locally integrable. To achieve this it suffices to observe that

$$\begin{aligned} E\int_0^\lambda |E[\delta(K_\alpha u'_\alpha)|U_\alpha = w]|d\alpha &= E\int_0^\lambda |E[\delta(K_\alpha u'_\alpha)|U_\alpha = w]|\frac{L_\alpha}{L_\alpha}d\alpha \\ &= E\int_0^\lambda |E[\delta(K_\alpha u'_\alpha)|U_\alpha]|\frac{1}{L_\alpha\circ U_\alpha}d\alpha \\ &= E\int_0^\lambda |E[\delta(K_\alpha u'_\alpha)|U_\alpha]|E[\rho(-\delta u_\alpha)|U_\alpha]d\alpha < \infty \end{aligned}$$

by hypothesis and by Theorem 2. Consequently we have the explicit expression for L_λ given as:

$$L_\lambda(w) = L_0 \exp\int_0^\lambda E[\delta(K_\alpha u'_\alpha)|U_\alpha = w]d\alpha\,.$$

□

Remark 1. *An important tool to control the hypothesis of Theorem 3 is the inequality of T. Carleman which says that (cf.* [3]*, Corollary XI.6.28)*

$$\|\det{}_2(I_H + A)(I_H + A)^{-1}\| \le \exp\frac{1}{2}\left(\|A\|_2^2 + 1\right)\,,$$

for any Hilbert-Schmidt operator A, where the left hand side is the operator norm, $\det_2(I_H + A)$ *denotes the modified Carleman-Fredholm determinant and $\|\cdot\|_2$ denotes the Hilbert-Schmidt norm. Let us remark that if A is a quasi-nilpotent operator, i.e., if the spectrum of A consists of zero only, then* $\det_2(I_H + A) = 1$*, hence in this case the Carleman inequality reads*

$$\|(I_H + A)^{-1}\| \le \exp\frac{1}{2}\left(\|A\|_2^2 + 1\right)\,.$$

This case happens when A is equal to the Sobolev derivative of some $u \in \mathbb{D}_{p,1}(H)$ whose drift $\dot{u}$ is adapted to the filtration $(\mathcal{F}_t,\, t \in [0,1])$,

From now on, for the sake of technical simplicity we shall assume that u_λ **is essentially bounded uniformly w.r.to λ.**

Proposition 1. *Let $F \in L^p(\mu)$ then the map $\lambda \to E[F|U_\lambda = w]$ is weakly continuous with values in $L^{p-}(\mu)$*[1]*.*

[1] $p-$ denotes any $p' < p$ and $q+$ any $q' > q$

Proof: First we have

$$\begin{aligned}\int_W |E[F|U_\lambda = w]|^p d\mu &= \int_W |E[F|U_\lambda = w]|^p \frac{L_\lambda}{L_\lambda} d\mu \\ &= \int_W |E[F|U_\lambda]|^{p-} \frac{1}{L_\lambda \circ U_\lambda} d\mu \\ &= \int_W |E[F|U_\lambda]|^p E[\rho(-\delta u_\lambda)|U_\lambda] d\mu < \infty\,,\end{aligned}$$

hence $E[F|U_\lambda = w] \in L^{p-}(\mu)$ for any $F \in L^p(\mu)$. Besides, for any $f \in C_b(W)$,

$$E[f \circ U_\lambda\, F] = E[f E[F|U_\lambda = w]\, L_\lambda]$$

therefore

$$|E[f \circ U_\lambda\, F]| \leq \|F\|_p \|f \circ U_\lambda\|_q \leq C_q \|F\|_p \|\|f\|_{q+}\,.$$

This relation, combined with the continuity of $\lambda \to f \circ U_\lambda$, due to the Lusin theorem, in L^q for any $f \in L^{q+}$, implies the weak continuity of the map $\lambda \to [F|U_\lambda = w]\, L_\lambda$ with values in $L^{p-}(\mu)$, since $\lambda \to L_\lambda$ and $\lambda \to (L_\lambda)^{-1}$ are almost surely and strongly continuous in $L^p(\mu)$, the claim follows. □

Theorem 4. *Assume that $F \in \mathbb{D}_{p,1}$ for some $p > 1$ and that*

$$E \int_0^\lambda |\delta(F K_\alpha u'_\alpha)| d\alpha < \infty$$

for any $\lambda > 0$, then $\lambda \to E[F|U_\lambda = w]$ is μ-a.s. absolutely continuous w.r. to the Lebesgue measure $d\lambda$ and the map $\lambda \to E[F|U_\lambda]$ is almost surely and hence L^p-continuous.

Proof: Using the same method as in the proof of Theorem 3, we obtain

$$\begin{aligned}\frac{d}{d\lambda} E[\theta \circ U_\lambda\, F] &= \frac{d}{d\lambda} E[\theta\, E[F|U_\lambda = w]\, L_\lambda] \\ &= E[\theta\, L_\lambda\, E[\delta(F\, K_\lambda u'_\lambda)|U_\lambda = w]]\end{aligned}$$

for any cylindrical function θ. By continuity w.r.to λ, we get

$$E\left[\theta\Big(L_\lambda E[F|U_\lambda = w] - L_0 E[F|U_0 = w]\Big)\right] = \int_0^\lambda E\left[\theta L_\alpha E[\delta(F K_\alpha u'_\alpha)|U_\alpha = w]\right] d\alpha\,.$$

By the hypothesis

$$E \int_0^\lambda |L_\alpha E[\delta(F K_\alpha u'_\alpha)|U_\alpha = w]| d\alpha < \infty$$

and since θ is an arbitrary cylindrical function, we obtain the identity

$$L_\lambda E[F|U_\lambda = w] - L_0 E[F|U_0 = w] = \int_0^\lambda L_\alpha\, E[\delta(F K_\alpha u'_\alpha)|U_\alpha = w] d\alpha$$

almost surely and this proves the first part of the theorem since $\lambda \to L_\lambda$ is already absolutely continuous and strictly positive. For the second part, we denote $E[F|U_\lambda]$ by $\hat{F}(\lambda)$ and we assume that $(\lambda_n, n \geq 1)$ tends to some λ, then there exists a sub-sequence $(\hat{F}(\lambda_{k_l}), l \geq 1)$ which converges weakly to some limit; but, from the first part of the proof, we know that $(E[F|U_{\lambda_{k_l}} = w], l \geq 1)$

converges almost surely to $E[F|U_\lambda = w]$ and by the uniform integrability, there is also strong convergence in $L^{p-}(\mu)$. Hence, for any cylindrical function G, we have

$$\begin{aligned} E[\hat{F}(\lambda_{k_l})\,G] &= E[E[F|U_{\lambda_{k_l}} = w]E[G|U_{\lambda_{k_l}} = w]L_{\lambda_{k_l}}] \\ &\to E[E[F|U_\lambda = w]E[G|U_\lambda = w]L_\lambda] \\ &= E[\hat{F}(\lambda)\,G]\,. \end{aligned}$$

Consequently, the map $\lambda \to \hat{F}(\lambda)$ is weakly continuous in L^p, therefore it is also strongly continuous. □

Remark: Another proof consists of remarking that

$$E[F|U_\lambda = w]|_{w=U_\lambda} = E[F|U_\lambda]$$

μ-a.s. and that $\lambda \to E[F|U_\lambda = w]$ is continuous a.s. and in L^{p-} from the first part of the proof and that $(L_\lambda, \lambda \in [a,b])$ is uniformly integrable. These observations, combined with the Lusin's theorem imply the continuity in $L^0(\mu)$ (i.e., in probability) of $\lambda \to E[F|U_\lambda]$ and the L^p-continuity follows.

We shall need some technical results, to begin with, let U_λ^τ denote the shift defined on W by

$$U_\lambda^\tau(w) = w + \int_0^{\cdot\wedge\tau} \dot{u}_\lambda(s)ds\,,$$

for $\tau \in [0,1]$. We shall denote by $L_\lambda(\tau)$ the Radon-Nikodym density

$$\frac{dU_\lambda^\tau\mu}{d\mu} = L_\lambda(\tau)\,.$$

Lemma 1. *We have the relation*

$$L_\lambda(\tau) = E[L_\lambda|\mathcal{F}_\tau]$$

almost surely.

Proof: Let f be an $\mathcal{F}_\tau$-measurable, positive, cylindrical function; then it is straightforward to see that $f\circ U_\lambda = f\circ U_\lambda^\tau$, hence

$$E[f\,L_\lambda] = E[f\circ U_\lambda] = E[f\circ U_\lambda^\tau] = E[f\,L_\lambda(\tau)]\,.$$

□

Lemma 2. *Let $\mathcal{U}_\lambda^\tau(t)$ be the sigma algebra generated by $\{U_\lambda^\tau(s);\, s \le t\}$. Then, we have*

$$E[f|\mathcal{U}_\lambda^\tau(1)] = E[f|U_\lambda^\tau]$$

for any positive, measurable function on W.

Proof: Here, of course the second conditional expectation is to be understood w.r. to the sigma algebra generated by the mapping U_λ^τ and once this point is fixed the claim is trivial. □

Proposition 2. *With the notations explained above, we have*

$$L_\lambda(\tau) = L_0(\tau)\exp\int_0^\lambda E[\delta\{(I_H+\nabla u_\alpha^\tau)^{-1}u_\alpha'^\tau\}|U_\alpha^\tau = w]d\alpha\,.$$

Moreover, the map $(\lambda,\tau)\to L_\lambda(\tau)$ is continuous on $\mathbb{R}\times[0,1]$ with values in $L^p(\mu)$ for any $p\geq 1$.

Proof: The first claim can be proved as we have done in the first part of the proof of Theorem 3. For the second part, let f be a positive, measurable function on W; we have

$$E[f\circ U_\lambda^\tau] = E[f\,L_\lambda(\tau)]\,.$$

If $(\tau_n,\lambda_n)\to(\tau,\lambda)$, from the Lusin theorem and the uniform integrability of the densities $(L_{\lambda_n}(\tau_n), n\geq 1)$, the sequence $(f\circ U_{\lambda_n}^{\tau_n}, n\geq 1)$ converges in probability to $f\circ U_\lambda^\tau$, hence, again by the uniform integrability, for any $q>1$ and $f\in L^q(\mu)$,

$$\lim_n E[f\,L_{\lambda_n}(\tau_n)] = E[f\,L_\lambda(\tau)]\,.$$

From Lemma 1, we have

$$\begin{aligned} E[L_{\lambda_n}(\tau_n)^2] &= E[L_{\lambda_n}(\tau_n)\,E[L_{\lambda_n}|\mathcal{F}_{\tau_n}]] \\ &= E[L_{\lambda_n}(\tau_n)\,L_{\lambda_n}]\,, \end{aligned}$$

since, from Theorem 3, $L_{\lambda_n}\to L_\lambda$ strongly in all L^p-spaces, it follows that $(\lambda,\tau)\to L_\lambda(\tau)$ is L^2-continuous, hence also L^p-continuous for any $p>1$. □

Proposition 3. *The mapping $(\lambda,\tau)\to L_\lambda(\tau)$ is a.s. continuous, moreover the map*

$$(\tau,w)\to(\lambda\to L_\lambda(\tau,w))$$

is a $C(\mathbb{R})$-valued continuous martingale and its restriction to compact intervals (of λ) is uniformly integrable.

Proof: Let us take the interval $\lambda\in[0,T]$, from Lemma 1 we have $L_\lambda(\tau) = E[L_\lambda|\mathcal{F}\tau]$, since $C([0,T])$ is a separable Banach space and since we are working with the completed Brownian filtration, the latter equality implies an a.s. continuous, $C([0,T])$-valued uniformly integrable martingale. □

Theorem 5. *Assume that*

$$E\int_0^\lambda\int_0^1\left(|\delta(\dot u_\alpha(s)K_\alpha u_\alpha')| + |\dot u_\alpha'(s)|^2\right)ds<\infty$$

for any $\lambda\geq 0$, then the map

$$\lambda\to E[\dot u_\lambda(t)|\mathcal{U}_\lambda(t)]$$

is continuous with values in $L_a^p(\mu,\,L^2([0,1],\mathbb{R}^d))$, $p\geq 1$.

Proof: Let $\xi\in L_a^\infty(\mu,H)$ be smooth and cylindrical, then, by similar calculations as in the proof of Theorem 4, we get

$$\begin{aligned} \frac{d}{d\lambda}E[(\xi\circ U_\lambda,u_\lambda)_H] &= \frac{d}{d\lambda}<\xi\circ U_\lambda,u_\lambda>=\frac{d}{d\lambda}<\xi\circ U_\lambda,\hat u_\lambda> \\ &= E\int_0^1\dot\xi_s L_\lambda(s)E\left[\delta(\dot u_\lambda(s)K_\lambda u_\lambda') + \dot u_\lambda'(s)|U_\lambda^s = w\right]ds\,, \end{aligned}$$

but the l.h.s. is equal to

$$E[(\nabla\xi\circ U_\lambda[u'_\lambda],u_\lambda)_H+(\xi\circ U_\lambda,u'_\lambda)_H]\,,$$

which is continuous w.r. to λ provided that ξ is smooth, and that $\lambda\to(u'_\lambda,u_\lambda)$ is continuous in L^p for $p\geq 2$. Consequently, we have the relation

$$<\xi\circ U_\lambda,u_\lambda>-<\xi\circ U_0,u_0>=E\int_0^\lambda\int_0^1\dot{\xi}_sL_\alpha(s)E\left[\delta(\dot{u}_\alpha(s)K_\alpha u'_\alpha)+\dot{u}'_\alpha(s)|U^s_\alpha=w\right]dsd\alpha$$

and the hypothesis implies that $\lambda\to L_\lambda(s)E[\dot{u}_\lambda(s)|U_\lambda^s=w]$ is μ-a.s. absolutely continuous w.r.to the Lebesgue measure $d\lambda$. Since $\lambda\to L_\lambda(s)$ is also a.s. absolutely continuous, it follows that $\lambda\to E[\dot{u}_\lambda(s)|U_\lambda^s=w]$ is a.s. absolutely continuous. Let us denote this disintegration as the kernel $N_\lambda(w,\dot{u}_\lambda(s))$, then

$$N_\lambda(U^s_\lambda(w),\dot{u}_\lambda(s))=E[\dot{u}_\lambda(s)|U^s_\lambda]$$

a.s. From the Lusin theorem, it follows that the map $\lambda\to N_\lambda(U^s_\lambda,\dot{u}_\lambda(s))$ is continuous with values in $L^0_a(\mu,L^2([0,1],\mathbb{R}^d))$ and the L^p-continuity follows from the dominated convergence theorem. □

Remark 2. *In the proof above we have the following result: assume that $\lambda\to f_\lambda$ is continuous in $L^0(\mu)$, then $\lambda\to f_\lambda\circ U_\lambda$ is also continuous in $L^0(\mu)$ provided that the family*

$$\left\{\frac{dU_\lambda\mu}{d\mu},\lambda\in[a,b]\right\}$$

is uniformly integrable for any compact interval $[a,b]$. To see this, it suffices to verify the sequential continuity; hence assume that $\lambda_n\to\lambda$, then we have

$$\begin{aligned}\mu\{|f_{\lambda_n}\circ U_{\lambda_n}-f_\lambda\circ U_\lambda|>c\}\quad&\leq\quad\mu\{|f_{\lambda_n}\circ U_{\lambda_n}-f_\lambda\circ U_{\lambda_n}|>c/2\}\\&\quad\quad+\mu\{|f_\lambda\circ U_{\lambda_n}-f_\lambda\circ U_{\lambda_n}|>c/2\}\,,\end{aligned}$$

but

$$\mu\{|f_{\lambda_n}\circ U_{\lambda_n}-f_\lambda\circ U_{\lambda_n}|>c/2\}=E[L_{\lambda_n}1_{\{|f_{\lambda_n}-f_\lambda|>c/2\}}]\to 0$$

by the uniform integrability of $(L_{\lambda_n},\,n\geq 1)$ and the continuity of $\lambda\to f_\lambda$. The second term tends also to zero by the standard use of Lusin theorem and again by the the uniform integrability of $(L_{\lambda_n},\,n\geq 1)$.

Corollary 1. *The map $\lambda\to E[\rho(-\delta u_\lambda)|U_\lambda]$ is continuous as an $L^p(\mu)$-valued map for any $p\geq 1$.*

Proof: We know that

$$E[[\rho(-\delta u_\lambda)|U_\lambda]=\frac{1}{L_\lambda\circ U_\lambda}\,.$$

□

Corollary 2. *Let $Z_\lambda(t)$ be the innovation process associated to U_λ, then*

$$\lambda\to\int_0^1E[\dot{u}_\lambda(s)|\mathcal{U}_\lambda(s)]dZ_\lambda(s)$$

is continuous as an $L^p(\mu)$-valued map for any $p\geq 1$.

Proof: We have

$$\log L_\lambda \circ U_\lambda = \int_0^1 E[\dot u_\lambda(s)|\mathcal{U}_\lambda(s)]dZ_\lambda(s) + \frac{1}{2}\int_0^1 |E[\dot u_\lambda(s)|\mathcal{U}_\lambda(s)]|^2 ds\,,$$

since the l.h.s. of this equality and the second term at the right are continuous, the first term at the right should be also continuous. □

Theorem 6. *Assume that*

$$E\int_0^\lambda |\delta\{\delta(K_\alpha u'_\alpha)K_\alpha u'_\alpha - K_\alpha \nabla u'_\alpha K_\alpha u'_\alpha + K_\alpha u''_\alpha\}|d\alpha < \infty$$

for any $\lambda \geq 0$. *Then the map*

$$\lambda \to \frac{d}{d\lambda}L_\lambda$$

is a.s. absolutely continuous w.r.to the Lebesgue measure $d\lambda$ *and we have*

$$\frac{d^2}{d\lambda^2}L_\lambda(w) = L_\lambda E[\delta D_\lambda | U_\lambda = w]\,,$$

where

$$D_\lambda = \delta(K_\lambda u'_\lambda)K_\lambda u'_\lambda - K_\lambda \nabla u'_\lambda K_\lambda u'_\lambda + K_\lambda u''_\lambda\,.$$

Proof: Let f be a smooth function on W, using the integration by parts formula as before, we get

$$\begin{aligned}
\frac{d^2}{d\lambda^2}E[f\circ U_\lambda] &= \frac{d}{d\lambda}E[f\circ U_\lambda\,\delta(K_\lambda u'_\lambda)]\\
&= E[(\nabla f\circ U_\lambda, u'_\lambda)_H \delta(K_\lambda u'_\lambda)]\\
&= E[(K_\lambda^\star \nabla(f\circ U_\lambda), u'_\lambda)_H \delta(K_\lambda u'_\lambda) + f\circ U_\lambda \delta(-K_\lambda \nabla u'_\lambda K_\lambda u'_\lambda + K_\lambda u''_\lambda)]\\
&= E\left[f\circ U_\lambda\left\{\delta(\delta(K_\lambda u'_\lambda)K_\lambda u'_\lambda) - \delta(K_\lambda \nabla u'_\lambda K_\lambda u'_\lambda) + \delta(K_\lambda u''_\lambda)\right\}\right].
\end{aligned}$$

Let us define the map D_λ as

$$D_\lambda = \delta(K_\lambda u'_\lambda)K_\lambda u'_\lambda - K_\lambda \nabla u'_\lambda K_\lambda u'_\lambda + K_\lambda u''_\lambda\,,$$

we have obtained then the following relation

$$\frac{d^2}{d\lambda^2}E[f\circ U_\lambda] = E[f\, L_\lambda\, E[\delta D_\lambda|U_\lambda = w]]$$

hence

$$< \frac{d}{d\lambda}L_\lambda, f > - < \frac{d}{d\lambda}L_\lambda, f > |_{\lambda=0} = \int_0^\lambda E[f\, L_\alpha\, E[\delta D_\alpha | U_\alpha = w]]d\alpha\,.$$

The hypothesis implies the existence of the strong (Bochner) integral and we conclude that

$$L'_\lambda - L'_0 = \int_0^\lambda L_\alpha E[\delta D_\alpha | U_\alpha = w]d\alpha$$

a.s. for any λ, where L'_λ denotes the derivative of L_λ w.r.to λ. □

Theorem 7. *Define the sequence of functionals inductively as*

$$\begin{aligned} D_\lambda^{(1)} &= D_\lambda \\ D_\lambda^{(2)} &= (\delta D_\lambda^{(1)}) K_\lambda u'_\lambda + \frac{d}{d\lambda} D_\lambda^{(1)} \\ &\ldots \\ D_\lambda^{(n)} &= (\delta D_\lambda^{(n-1)}) K_\lambda u'_\lambda + \frac{d}{d\lambda} D_\lambda^{(n-1)} . \end{aligned}$$

Assume that

$$E\int_0^\lambda |\delta D_\alpha^{(n)}| d\alpha < \infty$$

for any $n \geq 1$ and $\lambda \in \mathbb{R}$, then $\lambda \to L_\lambda$ is almost surely a C^∞-map and denoting by $L_\lambda^{(n)}$ its derivative of order $n \geq 1$, we have

$$L_\lambda^{(n+1)}(w) - L_0^{(n+1)}(w) = \int_0^\lambda L_\alpha E[\delta D_\alpha^{(n)} | U_\alpha = w] d\alpha .$$

4. Applications to the invertibility of adapted perturbations of identity

Let $u \in L_a^2(\mu, H)$, i.e., the space of square integrable, H-valued functionals whose Lebesgue density, denoted as $\dot{u}(t)$, is adapted to the filtration $(\mathcal{F}_t, t \in [0,1])$ dt-almost surely. A frequently asked question ire the conditions which imply the almost sure invertibility of the adapted perturbation of identity (API) $w \to U(w) = w + u(w)$. The next theorem gives such a condition:

Theorem 8. *Assume that $u \in L_a^2(\mu, H)$ with $E[\rho(-\delta u)] = 1$, let u_α be defined as $P_\alpha u$, where $P_\alpha = e^{-\alpha \mathcal{L}}$ denotes the Ornstein-Uhlenbeck semi-group on the Wiener space. If there exists a λ_0 such that*

$$\begin{aligned} & E\int_0^\lambda E[\rho(-\delta u_\alpha)|U_\lambda] \Big| E[\delta(K_\alpha u'_\alpha)|U_\alpha] \Big| d\alpha \\ & \qquad = E\int_0^\lambda E[\rho(-\delta u_\alpha)|U_\lambda] \Big| E[\delta((I_H + \nabla u_\alpha)^{-1} \mathcal{L} u_\alpha)|U_\alpha] \Big| d\alpha < \infty \end{aligned}$$

for $\lambda \leq \lambda_0$, then U is almost surely invertible. In particular the functional stochastic differential equation

$$\begin{aligned} dV_t(w) &= -\dot{u}(V_s(w), s \leq t) dt + dW_t \\ V_0 &= 0 \end{aligned}$$

has a unique strong solution.

Proof: Since u_α is an $H - C^\infty$-function, cf. [19], the API $U_\alpha = I_W + u_\alpha$ is a.s. invertible, cf.[20], Corollary 1. By the hypothesis and from Lemma 2 of [20], $(\rho(-\delta u_\alpha), \alpha \leq \lambda_0)$ is uniformly integrable. Let L_α and L be respectively the Radon-Nikodym derivatives of $U_\alpha \mu$ and $U\mu$ w.r. to μ. From Theorem 3,

$$L_\lambda(w) = L(w) \exp \int_0^\lambda E[\delta(K_\alpha u'_\alpha)|U_\alpha = w] d\alpha$$

for any $\lambda \leq \lambda_0$ and also that $\int_0^\lambda |E[\delta(K_\alpha u'_\alpha)|U_\alpha = w]| d\alpha < \infty$ almost surely. Consequently

$$L_\lambda - L = \left(\exp \int_0^\lambda E[\delta(K_\alpha u'_\alpha | U_\alpha = w] d\alpha - 1 \right) L \to 0$$

as $\lambda \to 0$, in probability (even in L^1). We claim that the set $(L_\alpha \log L_\alpha, \alpha \leq \lambda_0)$ is uniformly integrable. To see this let $A \in \mathcal{F}$, then

$$\begin{aligned}
E[1_A L_\alpha \log L_\alpha] &= E[1_A \circ U_\alpha \, \log L \circ U_\alpha] \\
&= -E[1_A \circ U_\alpha \, \log E[\rho(-\delta u_\alpha)|U_\alpha]] \\
&\leq -E[1_A \circ U_\alpha \, \log \rho(-\delta u_\alpha)] \\
&= E\left[1_A \circ U_\alpha \left(\delta u_\alpha + \frac{1}{2}|u_\alpha|_H^2\right)\right]
\end{aligned}$$

Since $(|u_\alpha|^2, \alpha \leq \lambda_0)$ is uniformly integrable, for any given $\varepsilon > 0$, there exists some $\gamma > 0$, such that $\sup_\alpha E[1_B |u_\alpha|^2] \leq \varepsilon$ as soon as $\mu(B) \leq \gamma$ and this happens uniformly w.r. to B, but as $(L_\alpha, \alpha \leq \lambda_0)$ is uniformly integrable, there exists a $\gamma_1 > 0$ such that, for any $A \in \mathcal{F}$, with $\mu(A) \leq \gamma_1$, we have $\mu(U_\alpha^{-1}(A)) \leq \gamma$ uniformly in α and we obtain $E[1_A \circ U_\alpha |u_\alpha|_H^2] \leq \varepsilon$ with such a choice of A. For the first term above we have

$$E[1_A \circ U_\alpha \delta u_\alpha] \leq E[1_A L_\alpha]^{1/2} \|u_\alpha\|_{L^2(\mu,H)} \leq \varepsilon$$

again by the same reasons. Hence we can conclude that

$$\lim_{\alpha \to 0} E[L_\alpha \log L_\alpha] = E[L \log L]\,.$$

Moreover, as shown in [15, 16], the invertibility of U_α is equivalent to

$$E[L_\alpha \log L_\alpha] = \frac{1}{2} E[|u_\alpha|_H^2] \to \frac{1}{2} E[|u|_H^2]\,,$$

therefore

$$E[L \log L] = \frac{1}{2} E[|u|_H^2]$$

which is a necessary and sufficient condition for the invertibility of U

□

In several applications we encounter a situation as follows: assume that $u : W \to H$ is a measurable map with the following property

$$|u(w+h) - u(w)|_H \leq c|h|_H$$

a.s., for any $h \in H$, where $0 < c < 1$ is a fixed constant, or equivalently an upper bound like $\|\nabla u\|_{op} \leq c$ where $\|\cdot\|_{op}$ denotes the operator norm. Combined with some exponential integrability of the Hilbert-Schmidt norm ∇u, one can prove the invertibility of $U = I_W + u$, cf. Chapter 3 of [19]. Note that the hypothesis $c < 1$ is indispensable because of the fixed-point techniques used to construct the inverse of U. However, using the techniques developed in this paper we can relax this rigidity of the theory:

Theorem 9. *Let $U_\lambda = I_W + \lambda u$ be an API (adapted perturbation of identity) with $u \in \mathbb{D}_{p,1}(H) \cap L^2(\mu, H)$, such that, for any $\lambda < 1$, U_λ is a.s. invertible. Assume that*

$$E \int_0^1 \rho(-\delta(\alpha u)) |E[\delta((I_H + \alpha \nabla u)^{-1} u)|U_\alpha]| d\alpha < \infty\,. \tag{4.4}$$

Then $U = U_1$ is also a.s. invertible.

Proof: Let $L = L_1$ be the Radon-Nikodym derivative of $U_1\mu$ w.r. to μ. It suffices to show that

$$E[L\log L] = \frac{1}{2}E[|u|_H^2]$$

which is an equivalent condition to the a.s. invertibility of U, cf. [16]. For this it suffices to show first that $(L_\lambda, \lambda < 1)$ converges in $L^0(\mu)$ to L, then that $(L_\lambda \log L_\lambda, \lambda < 1)$ is uniformly integrable. The first claim follows from the hypothesis (4.4) and the second claim can be proved exactly as in the proof of Theorem 8. □

5. Variational applications to entropy and estimation

In the estimation and information theories, one often encounters the problem of estimating the signal u_λ from the observation data generated by U_λ and then verifies the various properties of the mean square error w.r.to the signal to noise ratio, which is represented in our case with the parameter λ. Since we know that ([16])

$$E[L_\lambda \log L_\lambda] = \frac{1}{2}E\int_0^1 |E[\dot{u}_\lambda(s)|\mathcal{U}_\lambda(s)]|^2 ds\,,$$

the behavior of the mean square error is completely characterized by that of the relative entropy. Let θ denote the entropy of L_λ as a function of λ:

$$\theta(\lambda) = E[L_\lambda \log L_\lambda]\,.$$

From our results, it comes immediately that

$$\begin{aligned}
\frac{d\theta(\lambda)}{d\lambda} &= E[L'_\lambda \log L_\lambda] \\
&= E[L_\lambda\, E[\delta(K_\lambda u'_\lambda)|U_\lambda = w] \log L_\lambda] \\
&= E[E[\delta(K_\lambda u'_\lambda)|U_\lambda] \log L_\lambda \circ U_\lambda] \\
&= -E[\delta(K_\lambda u'_\lambda) \log E[\rho(-\delta u_\lambda)|U_\lambda]]\,.
\end{aligned}$$

Similarly

$$\begin{aligned}
\frac{d^2\theta(\lambda)}{d\lambda^2} &= E\left[L''_\lambda \log L_\lambda + (L'_\lambda)^2 \frac{1}{L_\lambda}\right] \\
&= E[L''_\lambda \log L_\lambda + L_\lambda\, E[\delta(K_\lambda u'_\lambda)|U_\lambda = w]^2] \\
&= E[E[\delta D_\lambda|U_\lambda = w] L_\lambda \log L_\lambda + +L_\lambda\, E[\delta(K_\lambda u'_\lambda)|U_\lambda = w]^2] \\
&= E[E[\delta D_\lambda|U_\lambda] \log L_\lambda \circ U_\lambda + E[\delta(K_\lambda u'_\lambda)|U_\lambda]^2]\,.
\end{aligned}$$

In particular we have

Theorem 10. *Assume that*

$$E\left[E[\delta D_\lambda|U_\lambda]\left(\int_0^1 E[\dot{u}_\lambda(s)|\mathcal{U}_\lambda(s)]dZ_\lambda(s) + \frac{1}{2}\int_0^1 |E[\dot{u}_\lambda(s)|\mathcal{U}_\lambda(s)]|^2 ds\right)\right] < E\left[E[\delta(K_\lambda u'_\lambda)|U_\lambda]^2\right]$$

for some $\lambda = \lambda_0 > 0$, *then there exists an* $\varepsilon > 0$ *such that the* **entropy is convex** *as a function of* λ *on the interval* $(\lambda_0 - \varepsilon, \lambda_0 + \varepsilon)$. *In particular, if* $u_0 = 0$, *then the same conclusion holds true on some* $(0, \varepsilon)$.

5.1. **Applications to the anticipative estimation.** In this section we study briefly the estimation of $\dot u_\lambda(t)$ with respect to the final filtration $\mathcal{U}_\lambda(1)=\sigma(U_\lambda)$.

Theorem 11. *Assume that*

$$E\int_0^\lambda L_\alpha|E[\dot u'_\alpha(s)+\delta(\dot u_\alpha(s)K_\alpha u'_\alpha)|U_\alpha]|^p d\alpha<\infty\,,$$

for a $p\geq 1$, then, dt-a.s., the map $\lambda\to L_\lambda E[\dot u_\lambda(t)|U_\lambda=x]$ and hence the map $\lambda\to E[\dot u_\lambda(t)|U_\lambda=x]$ are strongly differentiable in $L^p(\mu)$ for any $p\geq 1$ and we have

$$\frac{d}{d\lambda}E[\dot u_\lambda(t)|U_\lambda=x]=E[\dot u'_\lambda(t)+\delta(\dot u_\lambda(t)K_\lambda u'_\lambda)|U_\lambda=x]-E[\dot u_\lambda(t)|U_\lambda=x]E[\delta(K_\lambda u'_\lambda)|U_\lambda=x]$$

$d\mu\times dt$-a.s.

Proof: For a smooth function h on W, we have

$$\begin{aligned}\frac{d}{d\lambda}<E[\dot u_\lambda(t)|U_\lambda=x],h\,L_\lambda> &= \frac{d}{d\lambda}<E[\dot u_\lambda(t)|U_\lambda],h\circ U_\lambda>\\ &= E[\dot u'_\lambda(t)h\circ U_\lambda+\dot u_\lambda(t)(\nabla g\circ U_\lambda,u'_\lambda)_H]\\ &= E[E[\dot u'_\lambda(t)|U_\lambda]h\circ U_\lambda+h\circ U_\lambda\delta(\dot u_\lambda(t)K_\lambda u'_\lambda)]\\ &= E\left[hL_\lambda(x)\left(E[\dot u'_\lambda(t)|U_\lambda=x]+E[\delta(\dot u_\lambda(t)K_\lambda u'_\lambda)|U_\lambda=x]\right)\right]\,.\end{aligned}$$

The hypothesis implies that this weak derivative is in fact a strong one in $L^p(\mu)$, the formula follows by dividing both sides by L_λ and by the explicit form of L_λ given in Theorem 3. □

Using the formula of Theorem 11, we can study the behavior of the error of non-causal estimation of u_λ (denoted as NCE in the sequel) defined as

$$\begin{aligned}NCE &= E\int_0^1|\dot u_\lambda(s)-E[\dot u_\lambda(s)|\mathcal{U}_\lambda(1)]|^2ds\\ &= E\int_0^1|\dot u_\lambda(s)-E[\dot u_\lambda(s)|U_\lambda]|^2ds\end{aligned}$$

To do this we prove some technical results:

Lemma 3. *Assume that*

$$E\int_0^\lambda\int_0^1|\dot u''_\alpha(s)+\delta(\dot u'_\alpha(s)K_\alpha u'_\alpha)|^p ds d\alpha<\infty \tag{5.5}$$

for some $p>1$, for any $\lambda>0$, then the map

$$\lambda\to L_\lambda E[\dot u'_\lambda(s)|U_\lambda=x]$$

is strongly differentiable in $L^p_a(d\mu,L^2([0,1]))$, and its derivative is equal to

$$L_\lambda E[\dot u''_\lambda(s)+\delta(\dot u'_\lambda(s)K_\lambda u'_\lambda)|U_\lambda=x]$$

$ds\times d\mu$-a.s.

Proof: Let h be a cylindrical function on W, then, using, as before, the integration by parts formula, we get

$$\begin{aligned}\frac{d}{d\lambda}E[L_\lambda E[\dot{u}'_\lambda(s)|U_\lambda = x]\,h] &= \frac{d}{d\lambda}E[\dot{u}'_\lambda(s)\,h\circ U_\lambda]\\ &= E[\dot{u}''_\lambda(s)\,h\circ U_\lambda + h\circ U_\lambda\,\delta(\dot{u}'_\lambda(s)K_\lambda u'_\lambda)]\\ &= E\left[h\,L_\lambda\left(E[\dot{u}''_\lambda(s)+\delta(\dot{u}'_\lambda(s)K_\lambda u'_\lambda)|U_\lambda = x]\right)\right]\,.\end{aligned}$$

This proves that the weak derivative satisfies the claim, the fact that it coincides with the strong derivative follows from the hypothesis (5.5). □

Let us define the variance of the estimation as

$$\beta(\lambda,s) = E\left[|E[\dot{u}_\lambda(s)|\mathcal{U}_\lambda(1)]|^2\right]\,,$$

we shall calculate the first two derivatives of $\lambda\to\beta(\lambda,s)$ w.r.to λ in order to observe its variations. Using Lemma 3, we have immediately the first derivative as

$$\begin{aligned}\frac{d}{d\lambda}\beta(\lambda,s) &= E\Bigg[E[\dot{u}_\lambda(s)|U_\lambda = x]L_\lambda\\ &\quad\left(E[\dot{u}'_\lambda(s)+\delta(\dot{u}_\lambda(s)K_\lambda u'_\lambda)|U_\lambda = x] - \frac{1}{2}E[\dot{u}_\lambda(s)|U_\lambda = x]E[\delta(K_\lambda u'_\lambda)|U_\lambda = x]\right)\Bigg]\end{aligned} \tag{5.6}$$

The proof of the following lemma can be done exactly in the same manner as before, namely, by verifying first the weak differentaibility using cylindrical functions and then assuring that the hypothesis implies the existence of the strong derivative and it is left to the reader:

Lemma 4. *Assume that*

$$E\int_0^\lambda|\delta(\delta(K_\alpha u'_\alpha)K_\alpha u'_\alpha)+\delta(K_\alpha u''_\alpha - K_\alpha\nabla u'_\alpha K_\alpha u'_\alpha)|^p d\alpha<\infty\,,$$

for some $p\geq 1$. Then the map

$$\lambda\to L_\lambda E[\delta(K_\lambda u'_\lambda)|U_\lambda = x]$$

is strongly differentiable in $L^p(\mu)$ and we have

$$\begin{aligned}\frac{d}{d\lambda}(L_\lambda E[\delta(K_\lambda u'_\lambda)|U_\lambda = x]) &= L_\lambda E\left[\delta(\delta(K_\lambda u'_\lambda)K_\lambda u'_\lambda)|U_\lambda = x\right]\\ &\quad + L_\lambda E\left[\delta(K_\lambda u''_\lambda - K_\lambda\nabla u'_\lambda K_\lambda u'_\lambda)|U_\lambda = x\right]\,.\end{aligned}$$

Combining Lemma 3 and Lemma 4 and including the action of L_λ, we conclude that

$$\begin{aligned}
\beta''(\lambda) \quad = \quad & E\Big[E[\dot u''_\lambda + \delta(\dot u'_\lambda K_\lambda u'_\lambda)|U_\lambda]E[\dot u_\lambda(s)|U_\lambda]\Big] \\
& +E\Big[E[\dot u'_\lambda(s)|U_\lambda]\Big(E[\dot u'_\lambda(s) + \delta(\dot u_\lambda(s)K_\lambda u'_\lambda)|U_\lambda] \\
& \quad -E[\dot u_\lambda(s)|U_\lambda]E[\delta(K_\lambda u'_\lambda)|U_\lambda]\Big)\Big] \\
& +E\Big[E[\delta\left\{\dot u''_\lambda(s)K_\lambda u'_\lambda - \dot u'_\lambda(s)K_\lambda \nabla u'_\lambda K_\lambda u'_\lambda\right\} \\
& \quad +\delta\left\{\dot u_\lambda(s)K_\lambda u''_\lambda + \delta(\dot u_\lambda(s)K_\lambda u'\lambda)K_\lambda u'_\lambda\right\}|U_\lambda]E[\dot u_\lambda(s)|U_\lambda]\Big] \\
& +E\Big[E[\delta(\dot u_\lambda(s)K_\lambda u'_\lambda)|U_\lambda]\Big(E[\dot u'_\lambda(s) + \delta(\dot u_\lambda(s)K_\lambda u'_\lambda)|U_\lambda] \\
& \quad -E[\dot u_\lambda(s)|U_\lambda]E[\delta(K_\lambda u'_\lambda]|U_\lambda]\Big)\Big] \\
& -E\Big[E[\dot u_\lambda(s)|U_\lambda]\left(E[\dot u'_\lambda(s) + \delta(\dot u_\lambda(s)K_\lambda u'_\lambda)|U_\lambda] - E[\dot u_\lambda(s)|U_\lambda]E[\delta(K_\lambda u'_\lambda)|U_\lambda]\right)E[\delta(K_\lambda u'_\lambda)|U_\lambda]\Big] \\
& \quad -\frac{1}{2}E\Big[E[E[\dot u_\lambda(s)U_\lambda]^2\left\{E[\delta(\delta(K_\lambda u'_\lambda)K_\lambda u'_\lambda + K_\lambda u''_\lambda - K_\lambda\nabla u'_\lambda K_\lambda u'_\lambda)|U_\lambda]\right\}\Big]\,.
\end{aligned}$$

Assume now that $\lambda \to u_\lambda$ is linear, then a simple calculation shows that

$$\beta''(0) = E[|\dot u(s)|^2]\,,$$

hence the quadratic norm of the non-causal estimation of u, i.e., the function

$$\lambda \to E\int_0^1 |E[\dot u_\lambda(s)|\mathcal{U}_\lambda(1)]|^2 ds$$

is convex at some vicinity of $\lambda = 0$.

5.2. **Relations with Monge-Kantorovich measure transportation.** Since $L_\lambda \log L_\lambda \in L^1(\mu)$, it follows the existence of $\phi_\lambda \in \mathbb{D}_{2,1}$, which is 1-convex (cf. [4]) such that $(I_W + \nabla\phi_\lambda)\mu = L_\lambda \cdot \mu$ (i.e., the measure with density L_λ), cf. [5]. From the L^p-continuity of the map $\lambda \to L_\lambda$ and from the dual characterization of the Monge-Kantorovich problem, [21], we deduce the measurability of the transport potential ϕ_λ as a mapping of λ. Moreover there exists a non-causal Girsanov-like density Λ_λ such that

$$\Lambda_\lambda\, L_\lambda \circ T_\lambda = 1 \tag{5.7}$$

μ-a.s., where Λ_λ can be expressed as

$$\Lambda_\lambda = J(T_\lambda)\exp\left(-\frac{1}{2}|\nabla\phi_\lambda|_H^2\right)\,,$$

where $T_\lambda \to J(T_\lambda)$ is a log-concave, normalized determinant (cf.[6]) with values in $[0,1]$. Using the relation (5.7), we obtain another expression for the entropy:

$$\begin{aligned}
E[L_\lambda \log L_\lambda] &= E[\log L_\lambda \circ T_\lambda] \\
&= -E[\log \Lambda_\lambda] \\
&= E\left[-\log J(T_\lambda) + \frac{1}{2}|\nabla\phi_\lambda|_H^2\right]\,.
\end{aligned}$$

Consequently, we have

$$\begin{aligned}
\frac{1}{2}E\int_0^1 |E[\dot{u}_\lambda(s)\mid \mathcal{U}_\lambda(s)]|^2 ds &= E\left[-\log J(T_\lambda)+\frac{1}{2}|\nabla\phi_\lambda|_H^2\right]\\
&= E\left[-\log J(T_\lambda)\right]+\frac{1}{2}d_H^2(\mu, L_\lambda\cdot\mu)\,,
\end{aligned}$$

where $d_H(\mu, L_\lambda\cdot\mu)$ denotes the Wasserstein distance along the Cameron-Martin space between the probability measures μ and $L_\lambda\cdot\mu$. This result gives another explanation for the property remarked in [11] about the independence of the quadratic norm of the estimation from the filtrations with respect to which the causality notion is defined. Let us remark finally that if

$$d_H(\mu, L_\lambda\cdot\mu)=0$$

then $L_\lambda = 1$ μ-almost surely hence $E[\dot{u}_\lambda(s)\mid\mathcal{U}_\lambda(s)]=0$ $ds\times d\mu$-a.s. Let us note that such a case may happen without having $u_\lambda=0$ μ-a.s. As an example let us choose an API, say $K_\lambda = I_W + k_\lambda$ which is not almost surely invertible for any $\lambda\in(0,1]$. Assume that $E[\rho(-\delta k_\lambda)]=1$ for any λ. We have

$$\frac{dK_\lambda\mu}{d\mu}=\rho(-\delta m_\lambda)$$

for some $m_\lambda\in L_a^0(\mu,H)$, define $M_\lambda = I_W+m_\lambda$, then $U_\lambda = M_\lambda\circ K_\lambda$ is a Brownian motion and an API, hence (cf. [17]) it should be equal to its own innovation process and this is equivalent to say that $E[\dot{u}_\lambda(s)\mid\mathcal{U}_\lambda(s)]=0$ $ds\times d\mu$-a.s.

6. **Applications to Information Theory**

In this section we give first an extension of the results about the quadratic error in the additive nonlinear Gaussian model which extends the results of [1, 9, 10, 11] in the sense that we drop a basic assumption made implicitly or explicitly in these works; namely the conditional form of the signal is not an invertible perturbation of identity. Afterwards we study the variation of this quadratic error with respect to a parameter on whose depends the information channel in a reasonably smooth manner.

Throughout this section we shall suppose the existence of the signal in the following form:

$$U(w,m)=w+u(w,m)$$

where m runs in a measurable space $(M,\mathcal{M})$ governed with a measure ν and independent of the Wiener path w, later on we shall assume that the above signal is also parametrized with a scalar $\lambda\in\mathbb{R}$. We suppose also that, for each fixed m, $w\to U(w,m)$ is an adapted perturbation of identity with $E_\mu[\rho(-\delta u(\cdot,m))]=1$ and that

$$\int_0^1\int_{W\times M}|\dot{u}_s(w,m)|^2 ds d\nu d\mu<\infty\,.$$

In the sequel we shall denote the product measure $\mu\otimes\nu$ by γ and P will represent the image of γ under the map $(w,m)\to(U(w,m),m)$, moreover we shall denote by P_U the first marginal of P.

The following result is known in several different cases, cf. [1, 9, 10, 11], and we give its proof in the most general case:

Theorem 12. *Under the assumptions explained above the following relation between the mutual information $I(U,m)$ and the quadratic estimation error holds true:*

$$I(U,m)=\int_{W\times M}\log\frac{dP}{dP_U\otimes d\nu}dP=\frac{1}{2}E_\gamma\int_0^1\left(|E_\mu[\dot u_s(w,m)|\mathcal{U}_s(m)]|^2-|E_\gamma[\dot u_s|\mathcal{U}_s]|^2\right)ds\,,$$

where $(\mathcal{U}_s(m),\ s\in[0,1])$ is the filtration generated by the partial map $w\to U(w,m)$.

Proof. Let us note that the map $(s,w,m)\to E_\mu[f_s|\mathcal{U}_s(m)]$ is measurable for any positive, optional f. To proceed to the proof, remark first that

$$\frac{dP}{dP_U\otimes d\nu}=\frac{dP}{d\gamma}\frac{d\gamma}{dP_U\otimes d\nu}\tag{6.8}$$

$$\frac{d\gamma}{dP_U\otimes d\nu}=\frac{d\mu\otimes d\nu}{dP_U\otimes d\nu}=\left(\frac{dP_U}{d\mu}\right)^{-1}\tag{6.9}$$

since $P_U\sim\mu$. Think of $w\to U(w,m)$ as an API on the Wiener space for each fixed $m\in M$. The image of the Wiener measure μ under this map is absolutely continuous w.r. to μ; denote the corresponding density as $L(w,m)$. We have for any positive, measurable function f on $W\times M$

$$\begin{aligned}E_P[f]&=E_\gamma[f\circ U]\\&=\int_{W\times M}f(U(w,m),m)d\nu(m)d\mu(w)\\&=\int_M E_\mu\left[f\frac{dU(\cdot,m)\mu}{d\mu}\right]d\nu(m)\\&=E_\gamma[fL]\,,\end{aligned}$$

hence $(w,m)\to L(w,m)$ is the Radon-Nikodym density of P w.r. to γ. From [16] we have at once

$$E_\mu[L(\cdot,m)\log L(\cdot,m)]=\frac{1}{2}E_\mu\int_0^1|E_\mu[\dot u_s(\cdot,m)|\mathcal{U}_s(m)]|^2ds\,.$$

Calculation of $dP_U/d\mu$ is immediate:

$$\hat L=\frac{dP_U}{d\mu}(w)=\int_M L(w,m)d\nu(m).$$

Moreover from the Girsanov theorem, we have

$$E_\gamma[f\circ U\,\rho(-\delta u(\cdot,m))]=E_\gamma[f]$$

for any $f\in C_b(W)$. Denote by $\mathcal{U}_t$ the sigma algebra generated by $(U_s:\ s\le t)$ on $W\times M$. It is easy to see that the process $Z=(Z_t,t\in[0,1])$, defined by

$$Z_t=U_t(w,m)-\int_0^t E_\gamma[\dot u_s|\mathcal{U}_s]ds$$

is a γ-Brownian motion and any $(\mathcal{U}_t,\ t\in[0,1])$- local martingale w.r. to γ can be represented as a stochastic integral w.r. to the innovation process Z, cf. [7]. Let $\hat\rho$ denote

$$\hat\rho=\exp\left(-\int_0^1E_\gamma[\dot u_s|\mathcal{U}_s]dZ_s-\frac{1}{2}\int_0^1|E_\gamma[\dot u_s|\mathcal{U}_s]|^2ds\right)\tag{6.10}$$

Using again the Girsanov theorem we obtain the following equality

$$E_\gamma\left[f\circ U\hat\rho\right]=E_\gamma\left[f\circ U\rho(-\delta u(w,m))\right]$$

for any nice f. This result implies that

$$E_\gamma[\rho(-\delta u)|U] = \hat{\rho}$$

γ-almost surely. Besides, for nice f on W,

$$\begin{aligned} E_{P_U}[f] &= E_\gamma[f \circ U] = E_\gamma[fL] = E_\gamma[f\hat{L}] \\ &= E_\gamma[f \circ U \hat{L} \circ U \,\rho(-\delta u)] \\ &= E_\gamma[f \circ U \hat{L} \circ U \,\hat{\rho}] \end{aligned}$$

which implies that

$$\hat{L} \circ U \,\hat{\rho} = 1$$

γ-almost surely. We have calculated all the necessary ingredients to prove the claimed representation of the mutual information $I(U, m)$:

$$\begin{aligned} I(U,m) &= E_P\left[\log\left(\frac{dP}{d\gamma} \cdot \frac{d\gamma}{dP_U \otimes d\nu}\right)\right] \\ &= E_P\left[\log\frac{dP}{d\gamma} + \log\frac{d\gamma}{dP_U \otimes d\nu}\right] \\ &= E_\gamma\left[\frac{dP}{d\gamma}\log\frac{dP}{d\gamma}\right] - E_P\left[\log\frac{dP_U}{d\mu}\right] \\ &= E_\gamma\left[\frac{dP}{d\gamma}\log\frac{dP}{d\gamma}\right] - E_{P_U}\left[\log\frac{dP_U}{d\mu}\right] \\ &= E_\gamma\left[\frac{dP}{d\gamma}\log\frac{dP}{d\gamma}\right] - E_\mu\left[\frac{dP_U}{d\mu}\log\frac{dP_U}{d\mu}\right] \\ &= E_\gamma[L\log L] - E\gamma\left[\log\frac{dP_U}{d\mu}\circ U\right] \\ &= \frac{1}{2}E_\gamma\int_0^1 |E_\mu[\dot{u}_s(w,m)|\mathcal{U}_s(m)]|^2 ds - E_\gamma[-\log\hat{\rho}] \end{aligned}$$

and inserting the value of $\hat{\rho}$ given by the relation (6.10) completes the proof.

□

Remark: The similar results (cf. [1, 10, 11]) in the literature concern the case where the observation $w \to U(w, m)$ is invertible γ-almost surely, consequently the first term is reduced just to the half of the $L^2(\mu, H)$-norm of u (cf. [16]).
The following is a consequence of Bayes' lemma:

Lemma 5. *For any positive, measurable function g on $W \times M$, we have*

$$E_\gamma[g|U] = \frac{1}{\hat{L} \circ U}\left(\int_M L(x,m)E_\mu\Big[g \mid U(\cdot,m) = x\Big]d\nu(m)\right)_{x=U}$$

γ-almost surely. In particular

$$E_\gamma[g|U = x] = \frac{1}{\hat{L}(x)}\int_M L(x,m)E_\mu\Big[g \mid U(\cdot,m) = x\Big]d\nu(m)$$

P_U and μ-almost surely.

Proof. Let $f \in C_b(W)$ and let g be a positive, measurable function on $W \times M$. We have

$$\begin{aligned}
E_\gamma[g\, f \circ U] &= \int_M E_\mu[E_\mu[g \mid U(\cdot,m)]\, f \circ U(\cdot,m)] d\nu(m) \\
&= \int_M \int_W L(w,m)\, E_\mu[g \mid U(\cdot,m) = w]\, f(w) d\mu(w) d\nu(m) \\
&= \int_W f(w) \left(\int_M L(w,m) E_\mu[g \mid U(\cdot,m) = w]\, d\nu(m) \right) d\mu \\
&= \int_W \frac{\hat{L}(w)}{\hat{L}(w)} f(w) \left(\int_M L(w,m) E_\mu[g \mid U(\cdot,m) = w]\, d\nu(m) \right) d\mu \\
&= E_\gamma \left[\frac{1}{\hat{L} \circ U} f \circ U \left(\int_M L(w,m) E_\mu[g \mid U(\cdot,m) = w]\, d\nu(m) \right)_{w=U} \right]
\end{aligned}$$

□

From now on we return to the model U_λ parametrized with $\lambda \in \mathbb{R}$ and defined on the product space $W \times M$; namely we assume that

$$U_\lambda(w,m) = w + u_\lambda(w,m)$$

with the same independence hypothesis and the same regularity hypothesis of $\lambda \to u_\lambda$ where the only difference consists of replacement of the measure μ with the measure γ while defining the spaces $\mathbb{D}_{p,k}$.

Lemma 6. *Let $\hat{L}_\lambda(w)$ denote the Radon-Nikodym derivative of P_{U_λ} w.r. to μ. We have*

$$\hat{L}_\lambda(w) = \hat{L}_0(w) \exp \int_0^\lambda E_\gamma \Big[\delta(K_\alpha u'_\alpha) | U_\alpha = w \Big] d\alpha$$

μ-almost surely.

Proof. For any nice function f on W, we have

$$\frac{d}{d\lambda} E_\gamma[f \circ U_\lambda] = \frac{d}{d\lambda} E_\gamma[f\, L_\lambda] = \frac{d}{d\lambda} E_\mu[f\, \hat{L}_\lambda]\,.$$

On the other hand

$$\begin{aligned}
\frac{d}{d\lambda} E_\gamma[f \circ U_\lambda] &= E_\gamma[f \circ U_\lambda \delta(K_\lambda u'_\lambda)] \\
&= E_\gamma[f \circ U_\lambda E_\gamma[\delta(K_\lambda u'_\lambda) | U_\lambda]] \\
&= E_\gamma[f L_\lambda(x,m) E_\gamma[\delta(K_\lambda u'_\lambda) | U_\lambda = x]] \\
&= E_\mu[f \hat{L}_\lambda E_\gamma[\delta(K_\lambda u'_\lambda) | U_\lambda = x]]\,.
\end{aligned}$$

□

Remark: Note that we also have the following representation for $L_\lambda(w,m)$:

$$L_\lambda(w,m) = L_0(w,m) \exp \int_0^\lambda E_\mu \Big[\delta(K_\alpha u'_\alpha(\cdot,m)) | U_\alpha(\cdot,m) = w \Big] d\alpha$$

μ-a.s.

Lemma 7. *Let $\lambda \to \tau(\lambda)$ be defined as*

$$\tau(\lambda) = E_\gamma[\hat{L}_\lambda \log \hat{L}_\lambda]\,,$$

where $\hat{L}_\lambda(w) = \int_M L_\lambda(w,m)d\nu(m)$ *as before. We have*

$$\begin{aligned}\frac{d\tau(\lambda)}{d\lambda} &= E_\gamma\left[E_\gamma[\delta(K_\lambda u'_\lambda)|U_\lambda]\log \hat{L}_\lambda\circ U_\lambda\right]\\ &= E_\gamma\left[E_\gamma[\delta(K_\lambda u'_\lambda)|U_\lambda](-\log\hat{\rho}_\lambda)\right]\end{aligned}$$

where $\hat{\rho}_\lambda$ *is given by (6.10) as*

$$\hat{\rho}_\lambda = \exp\left(-\int_0^1 E_\gamma[\dot{u}_\lambda(s)|\mathcal{U}_\lambda(s)]dZ_\lambda(s) - \frac{1}{2}\int_0^1 |E_\gamma[\dot{u}_\lambda(s)|\mathcal{U}_\lambda(s)]|^2 ds\right).$$

Besides, we also have

$$\frac{d^2\tau(\lambda)}{d\lambda^2} = E_\gamma\left[E_\gamma[\delta D_\lambda|U_\lambda](-\log\hat{\rho}_\lambda) + E_\gamma[\delta(K_\lambda u'_\lambda)|U_\lambda]^2\right]$$

where

$$D_\lambda = \delta(K_\lambda u'_\lambda)K_\lambda u'_\lambda + \frac{d}{d\lambda}K_\lambda u'_\lambda.$$

Proof. The only thing that we need is the calculation of the second derivative of $\hat{L}_\lambda$: let f be a smooth function on W, then, from Lemma 5,

$$\begin{aligned}\frac{d^2}{d\lambda^2}E_\gamma[f\circ U_\lambda] &= \frac{d}{d\lambda}E_\gamma[f\circ U_\lambda\,\delta(K_\lambda u'_\lambda)]\\ &= E_\gamma\left[f\circ U_\gamma\delta\left(\delta(K_\lambda u'_\lambda)K_\lambda u'_\lambda + \frac{d}{d\lambda}(K_\lambda u'_\lambda)\right)\right]\\ &= E_\gamma[f\circ U_\gamma\,\delta D_\lambda]\\ &= E_\gamma[f(x)\,E_\gamma[\delta D_\lambda|U_\lambda = x]\,\hat{L}_\lambda(x)].\end{aligned}$$

□

As an immediate consequence we get

Corollary 3. *We have the following relation:*

$$\begin{aligned}\frac{d^2}{d\lambda^2}I(U_\lambda,m) &= E_\gamma\Big[E_\mu[\delta(D_\lambda(\cdot,m))|U_\lambda(m)](-\log E_\mu[\rho(-\delta u_\lambda(\cdot,m))|U_\lambda(m)])\\ &\quad + E_\mu[\delta(K_\lambda u'_\lambda(\cdot,m))|U_\lambda(m)]^2\Big]\\ &\quad - E_\gamma\Big[E_\gamma[\delta(D_\lambda)|U_\lambda](-\log\hat{\rho}_\lambda) + E_\gamma[\delta(K_\lambda u'_\lambda)|U_\lambda]^2\Big].\end{aligned}$$

References

[1] T. Duncan: "On the calculation of mutual information". SIAM, J. Appl. Math., vol. 19, 215-220, 1970.

[2] C. Dellacherie and P. A. Meyer: *Probabilités et Potentiel, Ch. I à IV*. Paris, Hermann, 1975.

[3] N. Dunford and J.T. Schwartz: *Linear Operators, Vol. 2*, New York, Interscience, 1967.

[4] D. Feyel and A.S. Üstünel. The notion of convexity and concavity on Wiener space. *Journal of Functional Analysis*, vol. 176,pp. 400-428, 2000.

[5] D. Feyel, A.S. Üstünel: Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space. *Probab. Theor. Relat. Fields*, **128**, no. 3, pp. 347–385, 2004.

[6] D. Feyel, A.S. Üstünel: "Log-concave measures". *TWMS J. Pure Appl. Math.* Vol. 1, No. 1, p. 92-105, 2010.

[7] M. Fujisaki, G. Kallianpur and H. Kunita:"Stochastic differential equations for the non linear filtering problem". Osaka J. Math., **9**, p. 19-40, 1972.

[8] D. Guo, S. Shamai and S . Verdú: "Mutual information and minimum mean-square error in Gaussian channels". IEEE transactions on Inf. Theory, Vol. 51, No. 4, p. 1261-1282, 2005.

[9] I.M. Gelfand and A.M. Yaglom: "Calculation of the amount of information about a random function contained in another such function". Usp. Mat. Nauk, vol. 12, 3-52, 1957 (transl. in Amer. Math. Soc. Transl., vol. 12, 199-246, 1959).
[10] T.T. Kadota, M. Zakai and J. Ziv: "Mutual information of the white Gaussian channel with and without feedback". IEEE Trans. Inf. Theory, vol. IT-17, no. 4, p. 368-371, 1971.
[11] E. Mayer-Wolf and M. Zakai: "Some relations between mutual information and estimation error in Wiener space". The Annals of Appl. Proba. Vol. 7, No.3, p. 1102-1116, 2007.
[12] M.S. Pinsker: *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, CA, 1964.
[13] A. S. Üstünel: *Introduction to Analysis on Wiener Space*. Lecture Notes in Math. Vol. 1610. Springer, 1995.
[14] A. S. Üstünel: *Analysis on Wiener Space and Applications*. http://arxiv.org/abs/1003.1649, 2010.
[15] A. S. Üstünel: "A necessary and sufficient condition for the invertibility of adapted perturbations of identity on the Wiener space". C.R. Acad. Sci. Paris, Ser. I, Vol.**346**, p. 897-900, 2008.
[16] A. S. Üstünel: "Entropy, invertibility and variational calculus of adapted shifts on Wiener space". J. Funct. Anal. 257 (2009), no. 11, 3655–3689.
[17] A. S. Üstünel: "Persistence of invertibility on the Wiener space". COSA, vol. 4, no. 2, p. 201-213, 2010.
[18] A. S. Üstünel and M. Zakai: "The construction of filtrations on abstract Wiener space". J. Funct. Anal. **143** , p. 10–32, 1997.
[19] A. S. Üstünel and M. Zakai: *Transformation of Measure on Wiener Space*. Springer Verlag, 1999.
[20] A. S. Üstünel and M. Zakai: "Sufficient conditions for the invertibility of adapted perturbations of identity on the Wiener space". Probab. Theory Relat. Fields, **139**, p. 207-234, 2007.
[21] C. Villani: *Topics in Optimal Transportation*. Graduate Series in Math., **58**. Amer. Math. Soc., 2003.
[22] M. Zakai: "On mutual information, likelihood ratios and estimation error for the additive Gaussian channel". IEEE Trans. Inform. Theory, **51**, 3017-3024, 2005.

A. S. Üstünel, Institut Telecom, Telecom ParisTech, LTCI CNRS Dépt. Infres,
46, rue Barrault, 75013, Paris, France
ustunel@enst.fr